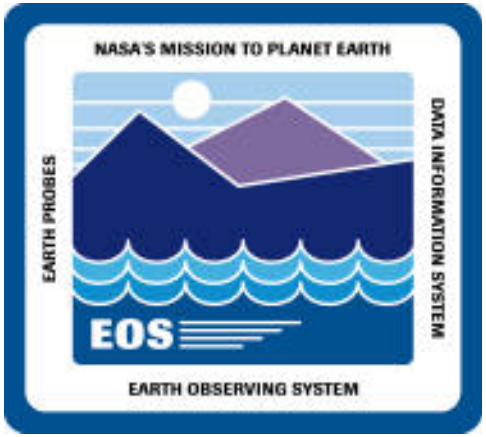


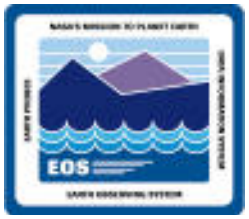
Processing Overview & Status

Simon Tackley

stackley@eos.hitc.com

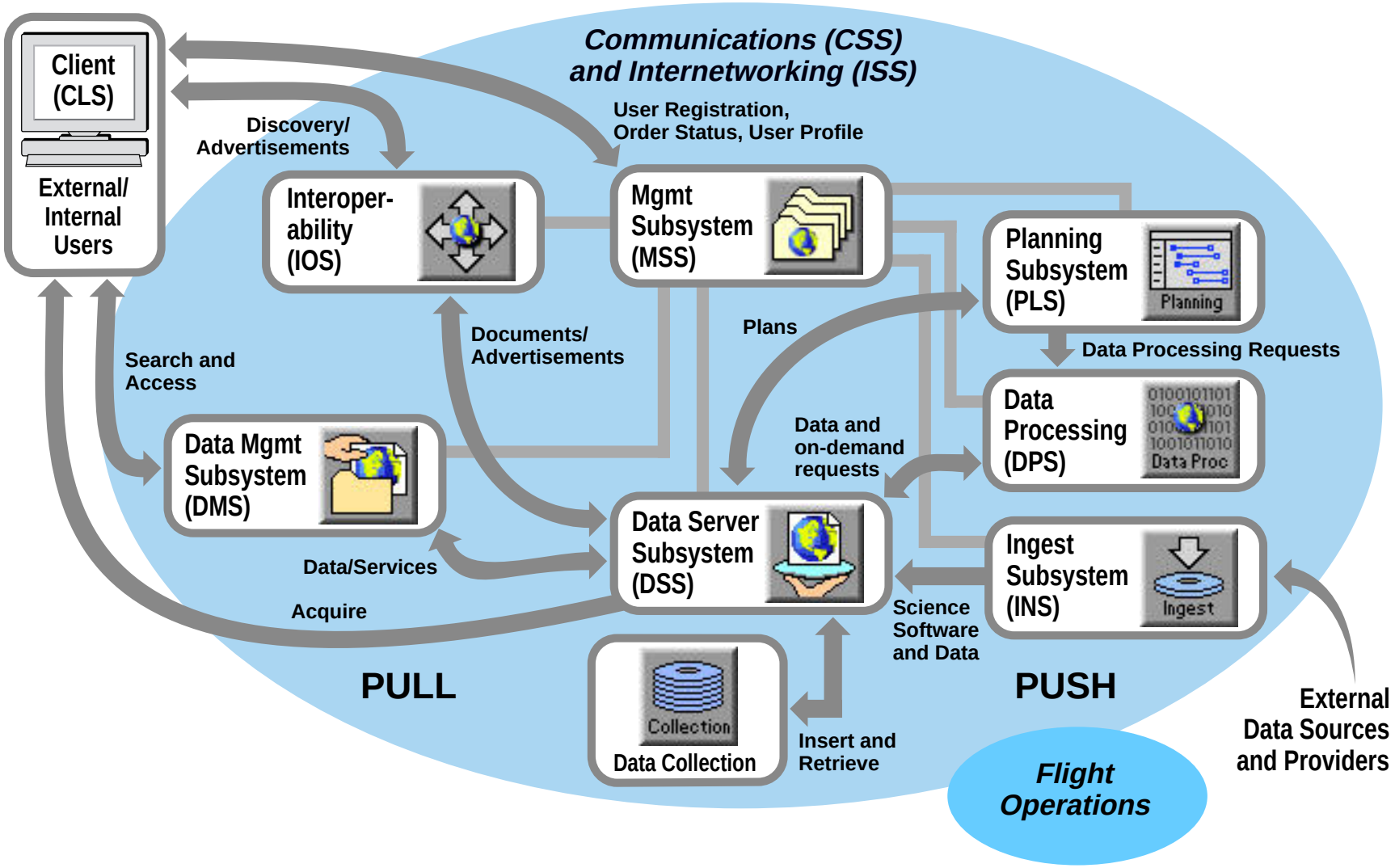
16 April 1996

Presentation Contents

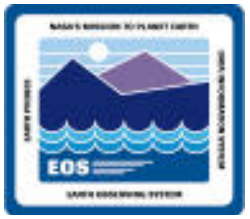


- ECS context
- Design drivers
- SW & HW Architecture
- COTS selections
- IDR & OWS workoffs
- Changes from IDR
- 305 Errata
- Prototypes and trades
- Candidate Release B functionality for the SDP Toolkit

ECS Context


Communications (CSS)
and Internetworking (ISS)
Client (CLS)
External/ Internal Users
Discovery/ Advertisements
User Registration, Order Status, User Profile
Interoper-ability (IOS)
Mgmt Subsystem (MSS)
Planning Subsystem (PLS)
Planning
Data Processing Requests
Search and Access
Documents/ Advertisements
Plans
Data Mgmt Subsystem (DMS)
Data and on-demand requests
Data Processing (DPS)
Data Proc
Data Server Subsystem (DSS)
Ingest Subsystem (INS)
Ingest
Data/Services
Acquire
Science Software and Data
PULL
PUSH
Collection
Data Collection
Insert and Retrieve
Flight Operations
External Data Sources and Providers

Release A Design Drivers by CSC



Algorithm Integration & Test

- Support SSIT
- Capture & store PGE Profile Information

QA

- Support in-line QA
- Support manual QA

Job Management

- Accept & process DPRs from PLS
- Cancel DPRs
- Update DPRs
- Interface to Autosys COTS

Data Preprocessing

- Reformat TRMM ancillary data
- Reformat SDPF L0 ancillary data

PDPS Database

- Persistent storage of all PDPS objects
- Primary interface between PLS & DPS

Autosys/AutoXpert COTS

- Queue DPRs with Priorities
- Execute jobs on remote machines
- Monitor production status
- Display production status

Execution Management

- Manage execution of PGEs
- Create SDP Toolkit PCF

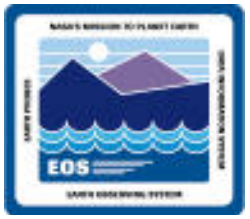
Data Management

- Maintain data map
- Interface to Data Server
- Stage data/PGEs from Data Server
- Destage data/PGEs to Data Server

Resource Management

- Maintain resource map
- Interface to MSS
- Allocate resources to DPRs
- Deallocate resources from DPRs

Release B Design Drivers by CSC



Algorithm Integration & Test

- Support Release B PGE information
- Create & store SSAP

QA

- Perform non-Science QA

Job Management

- Manage Predictive Staging
- Suspend/Resume DPRs
- Manage Autosys Load
- Monitor late DPR starts

Data Preprocessing

- Reformat EDOS FDS ancillary data

PDPS Database

Autosys/AutoXpert COTS

- Monitor higher throughputs
- Provide throughput to AHWGP Release B baseline

Execution Management

- Monitor PGE resource usage
- Manage PGE resource overruns
- PGE exit handling

Data Management

- Support avoidance of deadlock situations
- Support cancellation of staging/ destaging
- Support suspend/resume of staging/destaging
- Support cross-DAAC transfers

Resource Management

- Avoid deadlock situations
- Set aside resources for resource intensive PGEs

Release B Design Drivers

General Functional Description

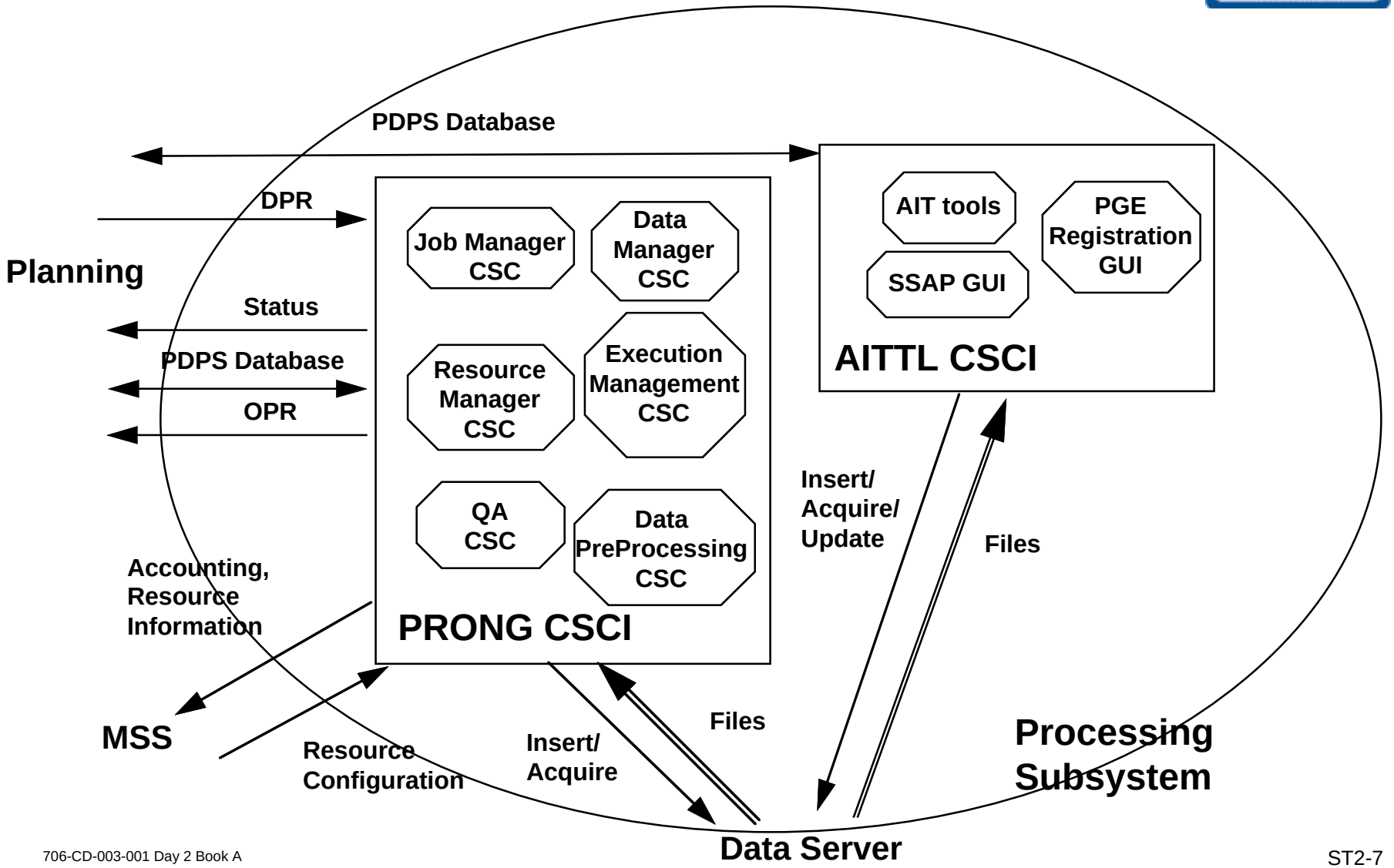
Performance Requirements

- meet AHWGP baseline throughput for Release B
- allow for upgrades in capacity

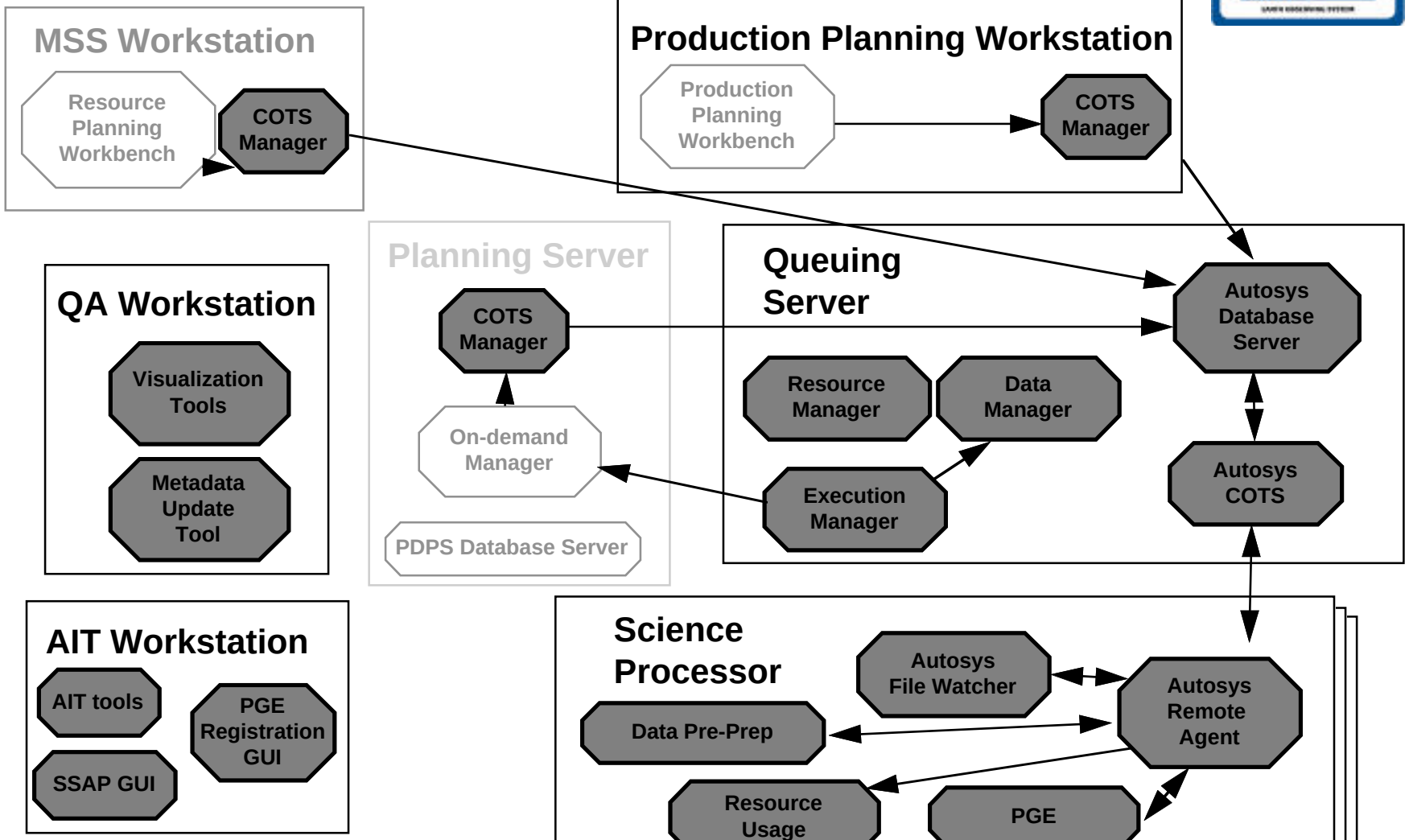
New Release B Features

- Predictive staging
  - staging data in advance of expected PGE execution
- Non-science QA
  - automated checking of metadata post PGE execution
- Data preprocessing
  - EDOS FDS data
- Resource management upgrades
  - avoid deadlock in support of predictive staging
  - handle resource intensive PGEs
- Suspend/Resume DPRs
- Automated exception handling
  - taking predetermined recovery paths on PGE failure
- Science Software Archive Package
  - storage of PGE Profile information in Data Server
- SSIT support for new PDPS features
  - Release B production rules
  - exception handling

Software Architecture Overview

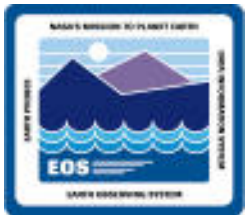


SW & HW Architecture



Most CSCs communicate with PDPS DB.
Not shown for readability reasons.

COTS Selections



COTS Inherited from Release A

- **Platinum Autosys**
- **Platinum AutoXpert**
- **EOSVIEW**
- **AIT Tools**
  - **Ghostview**
  - **Adobe Acrobat**
  - **SoftWindows**
  - **HTML Viewer**
  - **SPARCWorks (on the Sun)**
  - **CODEVision (on SGI)**
  - **CaseVision**
  - **ClearCase**
  - **Interactive Data Language**

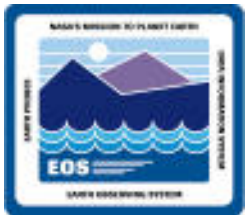
New COTS in Release B

- **none (some version upgrades)**

Status

- **no COTS evaluations planned**

IDR & OWS issues



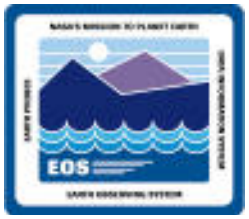
Status

- **IDR**
  - **All RIDs closed**
  - **Action Item closed**
  - **Issue closed**
  - **Comment closed**
- **Ops Workshop**
  - **All issues closed**

IDR and Ops Workshop Workoffs for Processing

Issue	Origin	Status	Resolution
Automatic notification of suspect granules	IDR RID #10	Closed	Updated QA design using subscriptions using metadata values
Rolling up Autosys error information	IDR AI #8	Closed	Execution management design updated to include logs GUI
Classification & response to PGE errors	IDR Issue #1	Closed	PDPS design updated to include actions based on PGE exit code
Handling of competing algorithms	IDR Comment #13	Closed	SSIT provides security & other features to support policy. PDPS design is policy neutral.
Is the subscription process capable of notifying based on QA flags?	OWS Issue 97	Closed	Updated Data Server Subscription Service design. QA design uses this feature. 305-CD-027-001 Section 4.5.6
ECS to update QA workflows to reflect current design	OWS Issue 99	Closed	Updated in DID605-2B Section 3.2.12.4

IDR and Ops Workshop Workoffs for Processing (cont.)

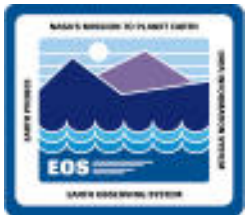


Issue	Origin	Status	Resolution
Provide criteria used by the DAAC to ensure integrity of data products	OWS Issue 98	Closed	Telecon 2/8/96 Included in Non-Science QA the design 305-CD-027-001 Section 4.3.7
CDR show resource allocation for adhoc, CAL/VAL production	IDR Action Item #2	Closed	Resource sizing is organized by 2 groups (AIT/On demand/Ad-hoc and standard production) only. Processing resources allocation design allows any resource configuration at the DAACs. 305-CD-027-002 Section 4.5.5
Discuss predictive staging with Sol Broder	IDR Action Item #5	Closed	Presented in PRONG CSCI presentation 305-CD-027-002 Section 4.5.2.1
Steve Marley to investigate if the production history is easily parsed for scripted retrieval	IDR Action Item #8	Closed	The production history can be parsed. It is incorporated into the ECS data model. 311-CD-002-004

IDR and Ops Workshop Workoffs for Processing (cont.)

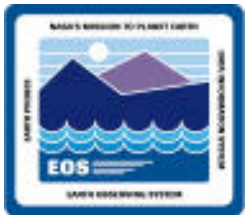
Issue	Origin	Status	Resolution
Better picture at CDR for Avg. and Peak performance, and how design can accommodate a rapid increase in processing	IDR Issue #38	Closed	305-CD-{014,015,016, 017}-001 Section 3.4.2.7.1
On demand production sizing should increase resolution by function for CDR	IDR Issue #48	Closed	305-CD-{014,015,016, 017}-001 Section 3.4.2.7.1
There is no requirement to process L0 expedited data	IDR Comment #15	Closed	L0 expedited data is received from EDOS at GSFC & LaRC. It does not require processing.
For CDR, investigate cross-DAAC resources for off-loading from a busy DAAC to a quiet DAAC. The best scenario would be for a non-critical task like "distributed reprocessing"	IDR Issue #54	Closed	The design does not preclude this occuring. However, there is a heavy performance hit from having to stage all the data across to the other DAAC. ECS is a *data-intensive* system.

Changes from IDR



- Predictive staging enhanced
- Resource management enhanced
  - event trace enhanced
  - enhanced anti-deadlock check
  - set-aside for resource hungry PGEs
- Automated failure handling
- Production Monitor higher level view
- Non-science QA
- Late start time monitoring
- Autosys throughput
- Incorporation of Key Mechanisms

Key Mechanisms



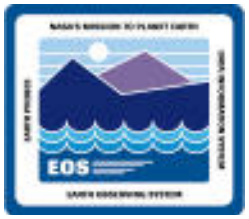
Mechanisms through which processes acquire or provide software services:

- **Process Framework (PF)**
  - ***Used in all CSCs except Data Preprocessing***
  - **Mechanism for incorporating infrastructures required to support distributed computing in the ECS environment**
  - **Provides process initialization and life cycle support**
  - **Sets parameters for naming/directory/security services**
  - **Interfaces to mode management, event logging, synchronous message passing**

Key Mechanisms (cont.)

- **Distributed Object Framework (DOF)**
  - ***Used in COTS manager CSC***
  - **Mechanism for creating remote objects and invoking remote methods**
  - **Provides naming, security, threads, time, and RPC services**
- **Universal Reference (UR)**
  - ***Used in all CSCs with Data Server interfaces***
  - **Mechanism for referencing system wide data and service objects**
  - **Provides externalize and internalize services for objects**

305 Errata



Failure of execution event trace **305-CD-027-002 4.5-27**
- added automatic exit status handling

Execution management object model **305-CD-027-002 4.3-4**
- added automatic exit status handling
- added deadlock avoidance and set-aside

PGE Registration object model **305-CD-027-002 6.3.3**
- added class for automatic exit status handling

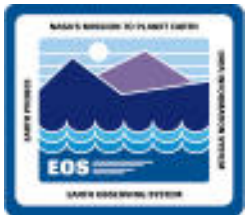
DpPrScheduler class definition **305-CD-027-002 4.4.67**
- PDL for predictive staging job added to for basic learning algorithm

Job Management object model **305-CD-027-002 4.3.2**
- added object for late start time monitoring
- added to for Autosys load handling / multiple Autosys'

Data Management object model **305-CD-027-002 4.3-3**
- GlCallback object should be EcSrAsyncRequest_C

Resource Management object model **305-CD-027-002 4.3-5**
- MsManager object should be EcAgEvent
- MsMsgCallback object should be EcAgManager

Prototypes and Trades



Completed :

- **Predictive Staging Trade study** **445-TP-004-001**

  Analysis of benefits and losses due to predictive staging techniques

  - **Design decisions**

    Predictive staging design

In Progress :

- **Autosys scalability benchmarking** **Poster session**

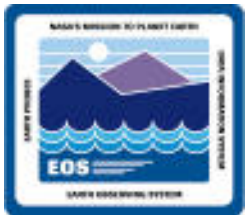
  Investigation of throughput of Autosys COTS under heavy job loads

  - **Design decisions**

    Processing Job Management design and hardware/software architecture complete

    Now investigating implementation issues (Sybase tuning etc.)

Candidate Release B Functionality for the SDP Toolkit



- **Improved Efficiency and Speed**
- **Support for PDPS handling of unusual PGE events**
- **HDF interface improvements based on user feedback to the Version 1 release**
- **User requested geolocation enhancements**
- **Enhanced product filenaming capability**